JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2002, **77,** 29–47 NUMBER 1 (JANUARY)

STABILITY OF FUNCTIONAL EQUIVALENCE AND STIMULUS EQUIVALENCE: EFFECTS OF BASELINE REVERSALS

OLIVER WIRTH AND PHILIP N. CHASE

WEST VIRGINIA UNIVERSITY

Functional equivalence and stimulus equivalence classes were established, reversed, and tested for stability with college students. Functional stimulus classes were established using a task in which students were trained to say nonsense words in the presence of arbitrarily assigned sets of symbols. Computer-controlled speech-recognition technology was used to record and analyze students' vocal responses for accuracy. After the establishment of stimulus classes was demonstrated with a transfer-of-function test, the effects of reversing selected baseline simple discriminations were assessed during an additional transfer-of-function test and a follow-up test that occurred several weeks later. With the same students, stimulus equivalence classes were established and demonstrated with computerized matching-to-sample procedures. The effects of reversing selected baseline conditional discriminations also were assessed during a postreversal equivalence test and a follow-up test. Both functional stimulus classes and stimulus equivalence were sensitive to contingency reversals, but the reversals with stimulus equivalence classes affected stimulus class organization whereas reversals with functional stimulus classes did not. Follow-up performances were largely consistent with the original baseline contingencies. The similarities and differences between stimulus equivalence and functional equivalence are related to the specific contingencies that select responding in the presence of the stimuli that form the classes.

Key words: stimulus classes, functional equivalence, stimulus equivalence, speech recognition, matching to sample, vocal behavior, humans

Novel behavior—behavior that does not have a history of direct reinforcement—has been shown to occur with respect to classes of stimuli that share no physical or perceptual features (e.g., shape or color). Such relations have created a special problem for behavioral accounts because, in part, the specific controlling variables are not easily identified. Despite considerable interest in this topic (see Zentall & Smeets, 1996, for a recent overview), the nature of those behavioral processes remains the subject of debate. Two experimental procedures that have been used to study such complex novel behavior with human and nonhuman subjects are those that establish stimulus equivalence classes and functional stimulus classes.

Stimuli that are arbitrarily assigned to a class are said to be related on the basis of stimulus equivalence when the properties of reflexivity, symmetry, and transitivity are demonstrated among the stimuli (Sidman et al., 1982; Sidman & Tailby, 1982). An additional relation, called equivalence or combined symmetry and transitivity, often is demonstrated as well. Stimulus equivalence classes typically are established using matching-to-sample (MTS) procedures in which conditional discriminations are arranged among arbitrarily assigned sets of stimuli.

Stimuli are said to be functionally equivalent when they occasion the same response and when they are functionally interchangeable (Goldiamond, 1962, 1966). For example, if the word ''car'' is spoken reliably in the presence of an actual car and a picture of a car, the stimuli are said to be functionally equivalent. In addition, functionally equivalent stimuli form a stimulus *class* if it can be shown that ''contingencies applied to one member of the class will affect other members of that class'' (Goldiamond, 1962, p. 303). That is, if the word ''car'' and the new word ''automobile'' are spoken and reinforced in the presence of an actual car, and then, without explicit training, the word ''automobile'' is spoken in the presence of a picture of a car, the actual car and picture of a car constitute a functional stimulus class (see also Catania, 1998). This generative aspect of functional stimulus classes has been described previously as mediated association or

Address all correspondence to Oliver Wirth, who is now at the National Institute for Occupational Safety and Health (NIOSH), 1095 Willowdale Rd. (MS 2027), Morgantown, West Virginia 26505 (E-mail: owirth@cdc.gov).

mediated transfer (Peters, 1935), mediated or semantic generalization (Cofer & Foley, 1942), and acquired equivalence (Goss, 1961). Functional stimulus classes typically are established and demonstrated through simple discriminations encompassed by the three-term contingency (e.g., Dube, McDonald, & McIlvane, 1991; Sidman, Wynne, Maguire, & Barnes, 1989; Vaughan, 1988).

Despite a growing interest in understanding the relations between functional equivalence and stimulus equivalence, the study of functional equivalence has not received comparable scrutiny. Furthermore, because stimulus equivalence and functional equivalence performances involve different behavioral units (conditional vs. simple discriminations) that are studied under disparate experimental procedures, they are not easily compared. As one step toward understanding the relation between these methods for establishing stimulus classes, studying functional equivalence under experimental manipulations that are commonly used to assess stimulus equivalence performances may uncover potentially important similarities or differences.

One issue that has been addressed with stimulus equivalence classes, but not with functional classes, is stimulus class stability, defined either as maintenance of performance over time or susceptibility to changes in defining contingencies. Assessment of stability can provide useful information about the variables that influence the maintenance or modifications of established or emergent discriminations. Previous reports have suggested that equivalence classes are stable over time (e.g., Rehfeldt & Hayes, 2000; Spradlin, Saunders, & Saunders, 1992), and that accurate performances can be maintained even after several months without intervening practice. Such stability seems important to the use of equivalence classes as a model for understanding complex stimulus control such as that found in verbal behavior.

Stability of equivalence classes also has been assessed by changing or reversing the prerequisite discriminations that defined the classes (Pilgrim, Chambers, & Galizio, 1995; Pilgrim & Galizio, 1990, 1995; R. Saunders, Saunders, Kirby, & Spradlin, 1988; Spradlin et al., 1992). In some of these studies, the effects of reversals have been found to differ across baseline, symmetry, and transitivity trial types. This outcome raises questions about whether equivalence relations should be conceptualized as integrated behavioral units or, instead, as a collection of independent stimulus-control relations. Reversing stimulus functions of established functional classes may serve similarly to reveal the extent to which prerequisite simple discriminations remain intact and whether the organization of the stimulus classes is altered. Furthermore, although previous studies have reversed baseline discriminations to test for functional class membership (e.g., Sidman et al., 1989; Smeets, 1994; Smeets, Barnes, & Luciano, 1995; Vaughan, 1988), no study has directly assessed the effects of reversals on the maintenance of functional stimulus classes over time.

In the present study, we directly assessed the stability of functional equivalence classes using reversal procedures and maintenance tests similar to those that have been used to study stimulus equivalence classes. We also attempted to replicate the findings on the stability of stimulus equivalence classes using reversal procedures and maintenance tests. To facilitate comparisons between functional and stimulus equivalence classes, we studied the stability of both with the same students. The procedures used to establish and demonstrate functional equivalence and stimulus equivalence also were equated as much as possible.

To establish functional stimulus classes, a speech-recognition procedure in which students speak nonsense words in the presence of stimuli was used to establish simple discriminations consistent with the formation of three three-member functional stimulus classes. This training was followed by a test of functional equivalence to ensure the establishment of three distinct functional stimulus classes. The baseline simple discriminations then were reversed for two target classes while those of a third class remained unchanged and served as controls. The impact of the baseline reversals on the pattern of performances was assessed in a postreversal test.

Using the same students, an MTS procedure featuring manual responses was used to establish conditional discriminations consistent with the formation of three three-member stimulus equivalence classes. After successful demonstrations of the properties of stimulus equivalence, two of three stimulus equivalence classes were targeted for baseline reversals while a third stimulus class was unaltered. The impact of the baseline reversals on the pattern of performances was assessed with tests of all possible discriminations among the stimuli. Follow-up tests were conducted after at least 4 weeks to determine if the passage of time affected the pattern of performances on both functional and stimulus equivalence classes.

METHOD

Participants

Eight undergraduate college students (7 females and 1 male, ages 18 to 29 years; hereafter, ''students'') were recruited to participate through a recruitment bulletin board located in the Psychology Department at West Virginia University. Students reported no experience with similar experiments. Each was paid 1¢ for each correct (i.e., class consistent) response in a block of trials in which an accuracy criterion (approximately 90%) was met. If the accuracy criterion was not met, no earnings were available for that block. In addition, students received a $1 bonus per session for attending all scheduled sessions.

Apparatus

Daily experimental sessions were conducted in a room (2.2 m by 1.8 m) equipped with a table, a chair, and the apparatus. The apparatus consisted of a microcomputer equipped with a 33-MHz 486 processor, 16 MB RAM, an IBM® M-ACPA sound card, a VGA color monitor, headphones, a VXI® headset microphone, and a keyboard on which the arrow (cursor) keys were positioned in an inverted-T configuration. The inverted-T configuration of the keys corresponded to the positions of stimuli displayed on the computer screen. Experimental events and data collection were controlled by C programming and Dragon VoiceTools™ 1.01 speech-recognition software (Wirth, Chase, & Munson, 2000). Throughout experimental sessions, the student wore a headset microphone and headphones for auditory feedback and to help mask extraneous noises.

Stimuli. Stimuli consisted of white symbols measuring approximately 2 cm by 2 cm on a computer screen (19 cm by 24 cm). They were displayed in one of four stimulus locations of blue squares (3 cm by 3 cm) on a black field. During speech-recognition trials, one stimulus appeared in a blue box at the center of the screen. During MTS trials, the blue squares were positioned in an inverted-T configuration. Sample stimuli appeared in the top square, and each of three comparison stimuli appeared in the squares positioned horizontally below the sample square.

Figure 1 shows the two sets of stimuli used. Each stimulus was assigned a number and a letter for descriptive purposes only. Numbers corresponded with experimenter-defined stimulus classes during the prereversal stages of the experiment and the letters A, B, and C designated stimulus class members. For example, A1, B1, and C1 were in the same class before the reversal procedures were implemented. On MTS trials, letters also designated sets of comparison stimuli. For example, B1, B2, and B3 served as comparison stimuli on the same trials.

Vocal responses used for speech-recognition trials also were assigned numbers that designated corresponding classes of stimuli. For example, Response 1 (GOX) was used to establish the functional stimulus class A1B1C1. Students S102, S104, S106, and S107 were trained with Set 1 (top panel of Figure 1) during reversals with stimulus equivalence classes and Set 2 (middle panel) during reversals with functional stimulus classes. The sets were reversed for S101, S103, S105, and S108.

General Procedure

By random assignment, students completed either the simple discrimination or MTS condition first. For S102, S103, S106, and S108, the effects of baseline reversals were assessed first on functional stimulus classes and then on stimulus equivalence classes; for S101, S104, S105, and S107, the order was reversed. Each session lasted approximately 50 min, and as many trial blocks as possible were conducted in a session. The postreversal test (described below) always was conducted on the day after the training criteria had been met to ensure that the amount of time between reversal training and testing was similar across students. Sessions were conducted 4 to 5 days per week across 2 to 3 weeks; students

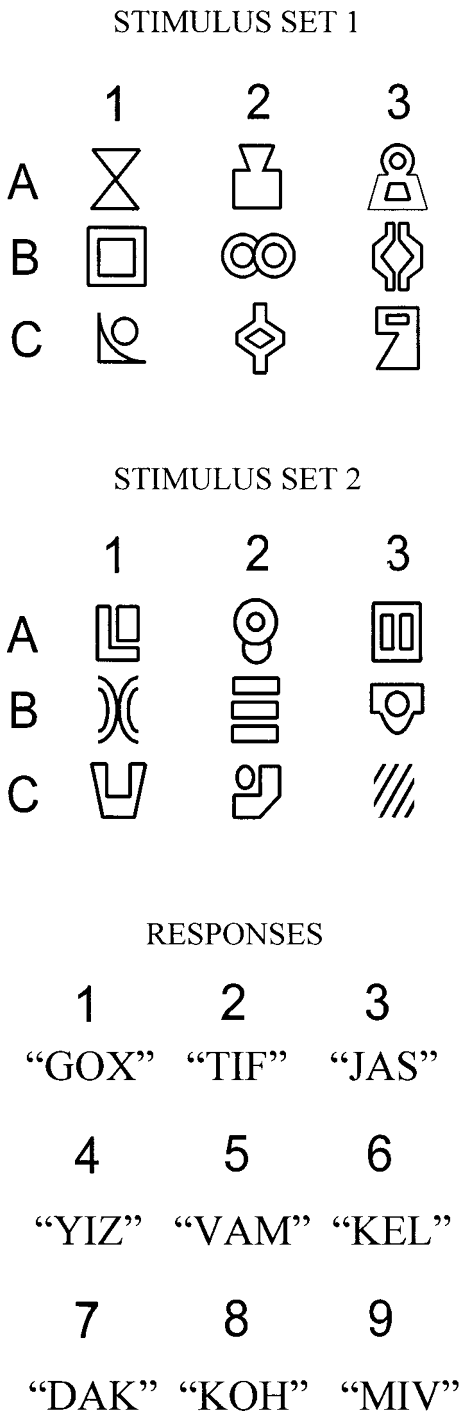

Fig. 1. Sets of stimuli and responses used to establish stimulus equivalence classes and functional stimulus classes.

returned after approximately 4 to 8 weeks for one follow-up session.

Baseline Reversals with Functional Stimulus Classes

Speech-recognition procedure. The speech-recognition apparatus and software were first adapted to accurately detect students' utterances of the first three nonsense words, GOX, TIF, and JAS (Responses R1, R2, and R3), according to the procedure described by Wirth et al. (2000). Speech-recognition training lasted approximately 5 min, after which each block of trials began with the following instructions displayed on the computer screen:

> During the next set of activities, your job will be to correctly name the symbols. Each trial will begin with the presentation of a symbol positioned inside a blue box at the center of the screen. If you know the correct name, say it out loud. If you don't know the correct name, wait and the correct name will appear on the screen—say it then. You can only earn money, however, if you make a correct response before it is displayed on the screen. Press "S" when you are ready to start.

Each trial began with the presentation of a stimulus. If no vocal response occurred within 20 s, students were prompted to make the correct response with the following displayed message: "Please say the correct name." Each correct response was followed by the presentation of the word "correct" at the bottom of the screen for 1 s and a 50-ms 2000-MHz tone. An incorrect response produced the word "incorrect" at the bottom of the screen for 1 s and a 50-ms 500-MHz tone.

Responses or other vocalizations that were not among the set of experimenter-defined words were considered neither correct nor incorrect. Instead, those responses were considered analogous to the "off-key" response that can occur with manual operanda. Following such responses, students were prompted to make another response with the following message displayed on the screen: "Not recognized—try again."

A response darkened the screen except for an empty blue box and initiated a 0-s to 2-s intertrial interval (ITI), which was varied across trials. To minimize responding prior to the presentation of the next stimulus, a vocal response during the ITI resulted in a 5-s delay to the presentation of the next stimulus. The variable and resetting features of the ITI were designed to increase the probability that students looked at the stimulus (Carlin, Wirth, & Chase, 1998). Key presses during trials had no programmed consequences. Stimuli were presented in a quasirandom sequence, with the restriction that no stimulus appeared on more than three consecutive trials. Earnings and percentage correct were displayed on the screen following the completion of each trial block, except during test trials.

Phase 1: Original baseline training. A graded delayed-prompt procedure (see Touchette, 1971) was used to minimize errors during training. During initial trial blocks when new discriminations were introduced, a delay between the presentation of a stimulus and the display of a written prompt for the correct response (e.g., ''Say JAS'') was increased from 2 s to 5 s and then was eliminated completely. Initial trial blocks began with a 2-s delay between the onset of a stimulus and the presentation of the response prompt approximately 2 cm below the stimulus. A response at any time after the onset of the stimulus initiated the appropriate feedback, but a correct response after the 2-s delay had elapsed initiated the additional message ''too slow,'' and no money was earned for that trial. When performance met a mastery criterion (22 of 24 trials correct for three consecutive trial blocks), the delay between the presentation of a stimulus and the correct response was increased to 5 s to facilitate responding before the prompt. When performance again met the criterion, performance was assessed in the absence of response prompts.

The original baseline discriminations were taught in four stages (see Phase 1 in Appendix A). Trial types involving the A, B, and C stimuli and corresponding responses hereafter are designated AR, BR, and CR. Initial trial blocks consisted of the three AR trial types (e.g., A1-1, A2-2, and A3-3), each presented eight times per block. After the accuracy criterion of 22 of 24 trials correct for three consecutive trial blocks had been met, the training of BR trial types (e.g., B1-1, B2-2, and B3-3) and then CR trial types (e.g., C1-1, C2-2, and C3-3) proceeded in the same manner. In the fourth stage of training, a mix of AR, BR, and CR trial types was presented four times in each block and until the standard accuracy criterion was met. Students were required to demonstrate accurate performance (i.e., 22 of 24 trials correct) for at least one trial block at each of four levels of reduced feedback (75%, 50%, 25%, and 0% of trials). For each block of trials in which feedback was reduced, students were informed about the differential feedback prior to the start of that block with the following additional note:

> We now want to see if you can remember what you have learned. During the next block of trials, you will not always be told if your response is correct or incorrect. The word ''correct'' or ''incorrect'' may not be displayed, and tones may not sound after every response. You will not be told how much money you have earned, but the experimenter still will keep a record of your earnings. Press ''S'' when you are ready to start.

Phase 2: Test of functional equivalence. A test for the interchangeability of stimulus functions (functional equivalence) was used to assess the establishment of functional equivalence among the class members (Goldiamond, 1962, 1966). This was accomplished by training a new response to one member of each class and then testing the remaining class members for a corresponding change in responding (i.e., transfer of function).

Following the establishment of original baseline discriminations, new responses YIZ, VAM, and KEL (Responses R4, R5, and R6) were introduced. The speech-recognition training procedure was repeated to add these new responses to the original three.

Following speech-recognition training, Responses R4, R5, and R6 were reinforced in the presence of Stimuli A1, A2, and A3, respectively, using the graded delayed-prompt procedure. Trial types (e.g., A1-4, A2-5, and A3-6) were presented eight times each for a total of 24 trials per block. Upon meeting the accuracy criterion of 22 of 24 trials correct for each of four levels of reduced feedback (75%, 50%, 25%, and 0%), test trials consisting of the A stimuli and each of six B and C stimuli (e.g., B1, B2, B3, C1, C2, and C3) were presented to test for transfer of function. Each stimulus was presented four times per block for a total of 36 trials. Test blocks were conducted in the absence of performance feedback, and they were repeated until accuracy was equal to or greater than 90%

or until accuracy across blocks was stable across a minimum of three trial blocks.

If transfer of function was not demonstrated, performance feedback was presented after each trial until responding was consistent with a transfer of function, and the test for transfer of function was repeated. First, the speech-recognition procedure was repeated to add three new responses, DAK, KOH, and MIV (Responses R7, R8, and R9), to the previous six. These responses then were reinforced in the presence of A stimuli in the manner described above. When performance met the accuracy criterion, test blocks that consisted of all A, B, and C stimuli were presented again to test for transfer of function. A student who failed to show transfer of function again on this second test was dropped from the experiment.

Phase 3: Baseline retraining. Following the tests for transfer of function that would suggest functional equivalence, students received additional training on the original baseline AR, BR, and CR trials with Responses R1, R2, and R3 as corresponding responses. This training ensured maintenance of accurate performance on the original baseline discriminations following tests for functional equivalence. These trial blocks were repeated until the accuracy criterion was met for a minimum of three successive trial blocks.

Phase 4: AR reversal training. In this phase, vocal responses corresponding to Stimuli A1 and A2 were reversed for two *target* classes, yielding the discriminations A1-2 and A2-1 (see Appendix A). That is, Response R2 was the correct response when Stimulus A1 was presented, and Response R1 was the correct response when A2 was presented. Because the class of A3, B3, and C3 served as the control, the discrimination A3-3 remained unchanged. In this phase, each stimulus (A1, A2, and A3) was presented eight times per block for a total of 24 trials. As in initial training phases, the graded delayed-prompt procedure was used to minimize errors during reversal training. Reversal training continued until the accuracy criterion was met. Thereafter, the proportion of trials with feedback was reduced across trial blocks to 75%, 50%, 25%, and 0%, with the criterion that accurate performance was demonstrated for at least one block at each level of feedback.

Phase 5: Postreversal test. Blocks that presented all A, B, and C stimuli were introduced next to assess the effects of reversing the original baseline discriminations (see Appendix A). In a quasirandom sequence, each stimulus was presented four times per block for a total of 36 trials. Trial blocks were conducted in the absence of feedback and prompts until stable performance was exhibited across a minimum of three trial blocks.

Baseline Reversals with Stimulus Equivalence Classes

MTS procedure. Each block of MTS trials began with the following instructions displayed on the computer screen:

> During the next set of activities, your job will be to select the correct symbol. Each trial will begin with the presentation of a symbol positioned inside a blue box at the center of the screen. A press on the up-arrow key will produce three additional blue boxes in which three different symbols will appear. If you know the correct response, press the left-, down-, or right-arrow key to ''select'' the left, middle, or right symbol, respectively. If you don't know the correct response, wait and the incorrect choices will disappear leaving only the correct choice on the screen. You can only earn money, however, if you select the correct symbol before the others disappear. Press ''S'' when you are ready to start.

Each trial began with the presentation of a sample stimulus. A press on the up-arrow key, which corresponded to the position of the sample stimulus, resulted in the presentation of three comparison stimuli. Previous research has suggested that the observing response facilitated acquisition of accurate MTS performance (Carlin et al., 1998). Upon presentation of the comparison stimuli, students could press either the left-, down-, or right-arrow key to select the left, middle, or right comparison stimuli, respectively. Positions of the comparison stimuli varied randomly across trials, with the restriction that stimuli could not appear in the same position on two consecutive trials. Selection of correct or incorrect comparisons produced feedback stimuli similar to those in speech-recognition trials. The selection of a comparison stimulus darkened the screen except for an empty blue sample box, and initiated a 0-s to 2-s ITI. To minimize responding prior to the presentation of the next sample stimulus, a press on any key during the ITI resulted in a 5-s delay to the presentation of the next sample stimulus. Sample stimuli were presented in a random sequence, with the restriction that no stimulus appeared on more than three consecutive trials. Earnings and percentage correct were displayed on the screen following the completion of each trial block, except during test phases.

Phase 1: Original baseline training. The original baseline conditional discriminations AB and BC were taught in three stages (see Appendix B). Initial trial blocks consisted of the three AB trial types (A1B1, A2B2, A3B3), each presented eight times per block. Upon meeting the accuracy criterion of 22 of 24 trials correct for three consecutive trial blocks, BC trial types (e.g., B1C1, B2C2, and B3C3) were presented until the accuracy criterion was met. In the third stage of training, blocks consisting of a mix of AB and BC trial types, each presented four times per block, were presented until the accuracy criterion was met.

A graded delayed-prompt procedure was used to minimize errors during training. During initial trial blocks when new conditional discriminations were introduced, selecting the sample stimulus presented three comparison stimuli and initiated a 2-s delay. When the 2-s delay elapsed, the two incorrect comparison stimuli disappeared, leaving only the correct comparison stimulus on the screen. A response selecting the correct (or class consistent) comparison stimulus before the 2-s delay had elapsed was considered correct, and appropriate feedback was presented to the student. Selecting the correct comparison stimulus after the 2-s delay had elapsed was considered correct, and appropriate feedback was presented; however, the additional message ''too slow'' was presented to the student and no money was earned for that trial. After performance met the standard accuracy criterion of at least 22 of 24 trials correct for three consecutive trial blocks, the prompt delay was increased to 5 s and then eliminated completely when performance again met the criterion.

Students were required to demonstrate accurate performance on blocks of mixed trial types with reduced feedback before advancing to the next phase. Accurate performance (at least 22 of 24 trials correct) was required for at least one trial block at each of four levels of reduced feedback (75%, 50%, 25%, and 0% of trials). For each block of trials with reduced feedback, students were informed about the differential feedback with the same note displayed in the previous condition.

Phase 2: Tests of stimulus equivalence. To determine whether training of the baseline conditional discriminations established the prerequisites for the formation of three three-member stimulus equivalence classes (i.e., A1B1C1, A2B2C2, and A3B3C3), blocks of trials that tested for the properties of stimulus equivalence were introduced next. These test blocks included trials that tested for reflexivity, symmetry, transitivity, or equivalence (i.e., combined transitivity and symmetry) interspersed among the original six original baseline conditional discriminations each presented four times (see Appendix B).

Reflexivity was tested first, followed by tests for symmetry, transitivity, and equivalence. (See Appendix B for the specific trial types presented.) Each test condition was in effect until accuracy was equal to or greater than 90% or was judged stable by visual inspection across a minimum of three blocks. Performance feedback was not presented during or after test blocks.

Phase 3: AB reversal training. Following the establishment and demonstration of three three-member stimulus equivalence classes, the original AB baseline conditional discriminations of the two target classes (e.g., A1B1 and A2B2) were reversed (see Appendix B). In this phase, when A1 and A2 were presented as sample stimuli, the correct comparison stimuli were B2 and B1, respectively, yielding the reversed discriminations A1B2 and A2B1. Because the class of A3, B3, and C3 served as the control, the original A3B3 baseline conditional discrimination remained unchanged. Trial blocks during the reversal phase consisted of all three AB trial types (A1B2, A2B1, and A3B3), each presented eight times per block for a total of 24 trials. As in original training phases, the graded delayed-prompt procedure was used to minimize errors during reversal training. Reversal training continued until the standard accuracy criterion was met. Thereafter, the proportion of trials with feedback was reduced across trial blocks to 75%, 50%, 25%, and 0% with the requirement that accurate performance be demon-

Table 1

Number of trial blocks conducted in each phase during the establishment and reversal of functional stimulus classes.

	Phase	Student					
		S102	S103	S104	S105	S106	S108
1.	Original baseline training						
	AR	9	10	10	10	10	13
	BR	9	9	9	10	10	10
	CR	9	10	10	10	9	10
	AR, BR, CR mix	7	7	7	7	7	8
2.	Test of functional equivalence						
	Function Change 1	14	14	14	14	15	14
	Transfer Test 1	3	3	1	1	3	3
	Criterion met?	No	No	No	No	No	Yes
	Function Change 2	14	14	15	14	13	
	Transfer Test 2	1	2	2	1	2	
	Criterion met?	Yes	Yes	Yes	Yes	Yes	
3.	Baseline retraining	3	3	3	3	7	4
4.	AR reversal training	13	14	14	15	18	14
5.	Postreversal test	3	3	3	3	4	3
6.	Follow-up test	3	3	3	3	3	3
	Days since postreversal test	35	36	31	32	70	33

strated for at least one block at each level of feedback.

Phase 4: Postreversal test. Portions of the tests from Phase 2 were repeated to assess responding across all possible trained and untrained trial types, including baseline conditional discriminations (see Appendix B). Specifically, the test assessed symmetry, transitivity, equivalence, and baseline relations. All trial types were intermingled in 72-trial blocks. No feedback was provided. Test blocks were repeated until performance was stable across a minimum of three trial blocks.

Follow-Up Tests

Students returned at least 4 weeks after the date of their last session of the experiment proper to assess whether the passage of time would affect the pattern of responding with respect to reversed stimulus classes. During a single follow-up session, the same tests of the earlier postreversal test phases were repeated (Phase 5 for functional classes and Phase 4 for equivalence classes). No feedback was provided. For each student, the tests were presented in the same order as in the main part of the experiment.

RESULTS

Results are presented for 6 of 8 students who successfully demonstrated transfer of function during the tests for functional equivalence. Because S101 and S107 did not demonstrate transfer, they were excluded from further analyses.

Baseline Reversals with Functional Stimulus Classes

Baseline training and transfer test. Table 1 shows the number of trial blocks required by each student to meet the criterion in all phases. All students acquired the AR, BR, and CR baseline simple discriminations and required similar amounts of training during the original baseline training phase. Performance met criterion after 9 to 13 blocks with each trial type and seven to eight mixed blocks.

This baseline training, however, was not sufficient to produce functional stimulus classes. Only 1 student demonstrated transfer of function in the first test. Other students required a second function change followed by another transfer test before transfer of function was demonstrated.

Reversal training and postreversal test. Table 1 also shows that students required between 13 and 18 trial blocks to meet the accuracy criterion for responding on the reversed AR trials, and then three to four trial blocks to demonstrate consistent patterns of responding on the postreversal test.

Figure 2 shows the pattern of simple-discrimination performances across each trial

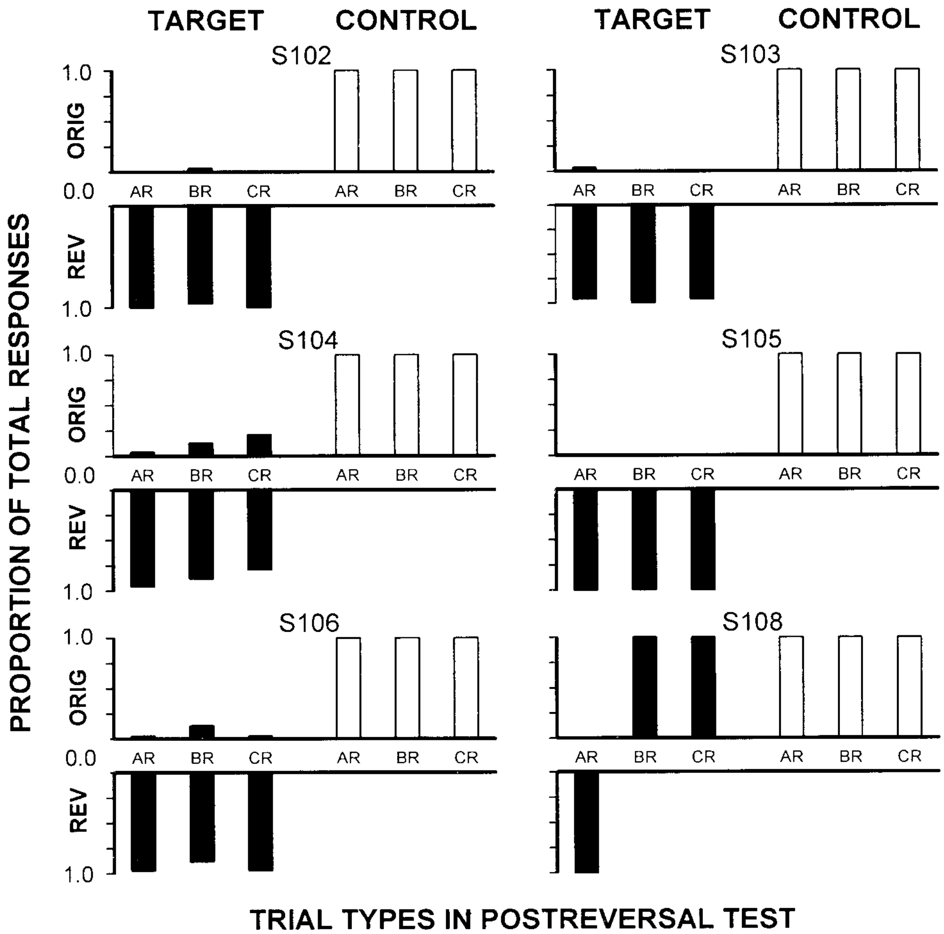

Fig. 2. Proportion of responses of each speech-recognition trial type consistent with the original baseline training (ORIG) or AR reversal training (REV) aggregated across all trial blocks of the postreversal test. Response patterns are shown for the combined trial types of the two stimulus classes targeted for AR reversals (TARGET) and for the third control class for which the AR baseline trials were not reversed (CONTROL).

type of the postreversal test. The proportion of total responses consistent with either original baseline training or AR reversal training is depicted by the direction of the bars. Bars directed upward depict a pattern consistent with the original baseline training; bars directed downward depict a pattern consistent with AR reversal training. Because performances with respect to the two stimulus classes targeted for baseline reversals were similar, these performances were aggregated.

As Figure 2 shows, the patterns of all students' responding across trial types of the control class remained consistent with the original baseline training. Conversely, for most students the patterns of responding across AR, BR, and CR trials of the two stimulus classes targeted for reversals were consistent in general with the reversal training. That is, Stimuli A1, B1, and C1 now occasioned Response R2, and Stimuli A2, B2, and C2 occasioned Response R1. Responding of S108 showed complete reversal only on AR trials.

Baseline Reversals with Stimulus Equivalence Classes

Training and equivalence tests. Table 2 shows the number of trial blocks required by each student to meet the criterion in all phases. Students acquired the original AB and BC baseline conditional discriminations after 9 to 12 blocks with each baseline trial type and 7 to 10 mixed blocks.

Table 2

Number of trial blocks conducted in each phase during the establishment and reversal of stimulus equivalence classes.

	Phase	Student					
		S102	S103	S104	S105	S106	S108
1.	Original baseline training						
	AB	10	10	12	11	10	11
	BC	10	9	11	9	11	10
	AB, BC mix	7	8	10	7	8	7
2.	Test of stimulus equivalence						
	Reflexivity	1	2	8	1	1	1
	Symmetry	1	1	2	1	1	1
	Transitivity	1	1	5	1	1	6
	Equivalence	1	1	1	1	1	2
3.	AB reversal training (first/second exposure)	14/2[a]	14/2[a]	13	13	14/1[a]	14
4.	Postreversal test (first/second exposure)	3/3[a]	3/3[a]	4	3	4/2[a]	3
5.	Follow-up test	3	3	3	3	3	3
	Days since postreversal test	31	29	37	42	67	31

[a] Because responding on AB trials during the first postreversal tests did not reverse for these subjects, AB reversal training and postreversal tests were repeated; see text for details.

On tests of reflexivity, symmetry, transitivity, and equivalence, most students required only one trial block to demonstrate criterion performance. Two students (S104 and S108), however, required several presentations of some test blocks before criterion performance was demonstrated. Eventually, all students demonstrated the formation of three three-member stimulus equivalence classes (i.e., A1B1C1, A2B2C2, and A3B3C3) prior to AB reversal training.

Reversal training and postreversal test. During reversal training, all students required 13 or 14 trial blocks to reverse their responding on the AB trials corresponding to the first two stimulus equivalence classes targeted for reversal. Thus, students now selected B2 in the presence of A1 and B1 in the presence of A2. As expected, responding remained unaltered on the third AB trial type (A3B3).

Blocks of trials that tested for all possible conditional discriminations were presented during the postreversal test to assess the effects of AB baseline reversals on the organization of stimulus equivalence classes (see Phase 4 in Appendix B). Table 2 shows that students required three to four trial blocks to demonstrate stable patterns of responding on the postreversal test. S102, S103, and S106, however, consistently selected B1 in the presence of A1 and B2 in the presence of A2 according to the original baseline training, despite having reversed these conditional discriminations in the reversal training of the previous phase. For these students, symmetry, transitivity, and equivalence also remained consistent with the original class structures. Thus, the series of AB reversal training and the postreversal tests were repeated to continue to use these students for further analyses. The number of additional trial blocks required by S102, S103, and S106 to show the reversed classes were three, three, and two, respectively.

Figure 3 shows the pattern of MTS performance on each trial type after AB reversal training. As in Figure 2, the proportion of total responses consistent with either original baseline training or AB reversal training is depicted by the direction of the bars. Trial types that were presented during the postreversal test included the baseline (AB and BC), symmetry (BA and CB), transitivity (AC), and equivalence (CA) trials.

The right panels of Figure 3 show that the pattern of responding across all trial types of the control class remained consistent with the original baseline training. Only S106 showed decrements in performance. On trials of the two stimulus classes targeted for AB reversals, all students except S103 reversed their responding on AB baseline trials, as expected, and on the corresponding BA symmetry trials (i.e., B1A2 and B2A1). Because the original

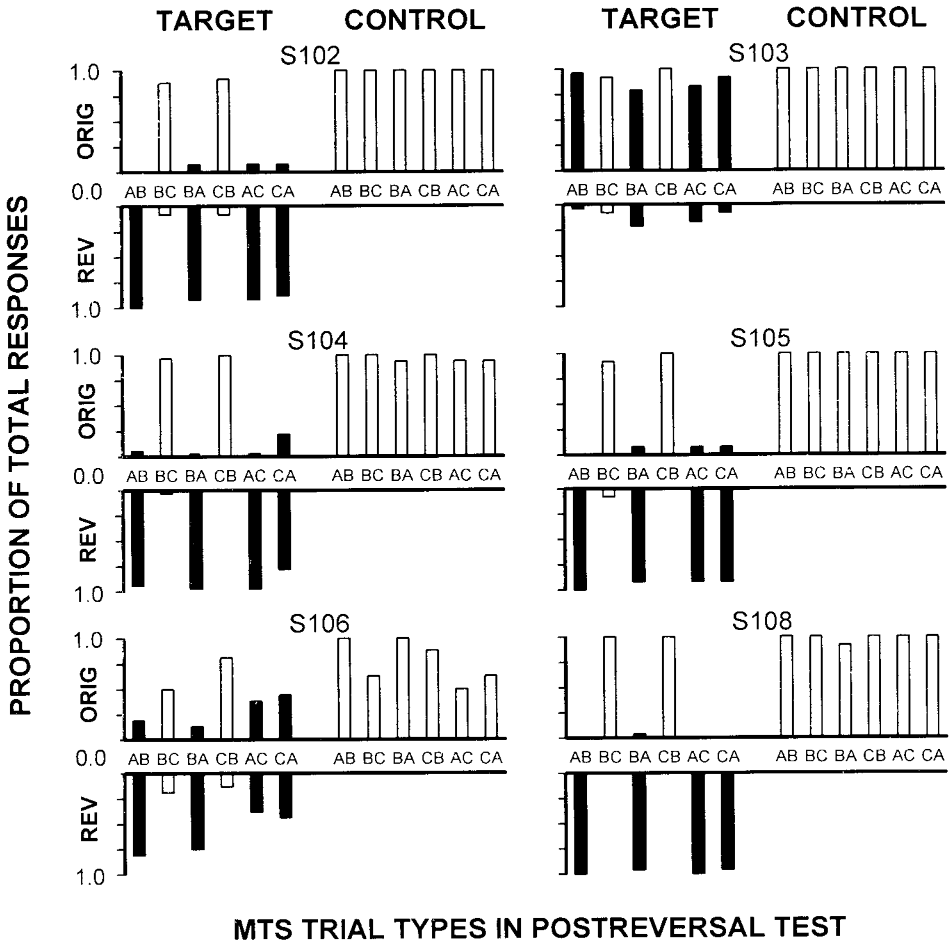

Fig. 3. Proportion of responses on each MTS trial type consistent with the original baseline training (ORIG) or AB reversal training (REV) aggregated across all blocks of the postreversal test. For S102, S103, and S106, performance is shown for the second postreversal test only. Response patterns are shown for the combined trial types of the two stimulus classes targeted for AB reversals (TARGET) and the third control class for which the AB baseline trials were not reversed (CONTROL). Filled bars depict trial types predicted to reverse; open bars depict responding predicted to remain consistent with original training.

BC baseline trials were not altered, responding on those trials and corresponding CB symmetry trials remained consistent with original training, as shown by most students. Only S106 responded inconsistently on BC and CB trial types. In general, the pattern of responding on AC transitivity and CA equivalence trial types also was consistent with the AB reversal training. This pattern was clear for S102, S104, S105, and S108, and to a lesser extent (approximately 30% to 50% of trials) for S106. Despite S103's second exposure to the series of AB reversal training and testing, her responding remained consistent with original baseline training across all trial types. In general, the results shown in Figure 3 demonstrate that, for most students, even those who initially resisted reversing classes (S102, S103, and S106), a complete reorganization occurred with two of the three stimulus classes (A2B1C1 and A1B2C2).

Follow-Up Tests

Functional stimulus classes. Figure 4 shows the patterns of responding on trial types of the previously established functional stimulus classes during the follow-up tests. The patterns across trials of the target classes were

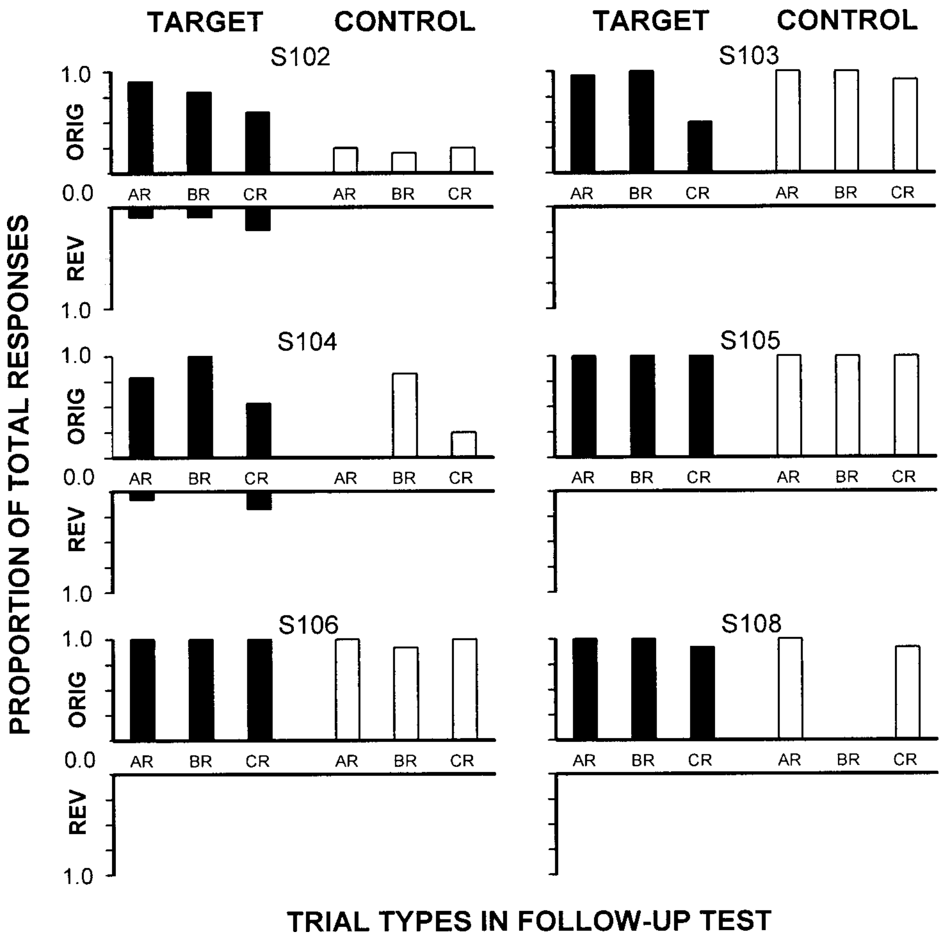

Fig. 4. Proportion of responses on each speech-recognition trial type consistent with original baseline training (ORIG) or AR reversal training (REV) aggregated across all trial blocks of the follow-up test. All other details as in Figure 2.

largely consistent with the original baseline training. That is, Responses R1 and R2 occurred in the presence of stimuli of the target classes and R3 occurred in the presence of stimuli of the control class. S102, S104, and S108, however, showed disruption in performance on one or more trial types of the control class.

Stimulus equivalence classes. Figure 5 shows the patterns of responding on trial types of the previously established stimulus equivalence classes during the follow-up test. A comparison of the patterns of responding during the postreversal and follow-up tests (Figures 3 and 5) shows that 4 of 6 students changed from a pattern previously consistent with the reversal training to one consistent with the original baseline training. S106 and S108, however, showed marked disruption in responding across both target and control classes. Responding of these students was not consistent with original baseline training or AB reversal training, even on trials of the control class for which baseline trials were not explicitly altered.

DISCUSSION

Goldiamond's (1962) definition of functional equivalence suggests that training many responses to the same stimulus set will ultimately allow a new response, trained to one stimulus in the set, to be occasioned by the other stimuli in the set without training. This definition appears to be consistent with the acquisition of many everyday verbal rela-

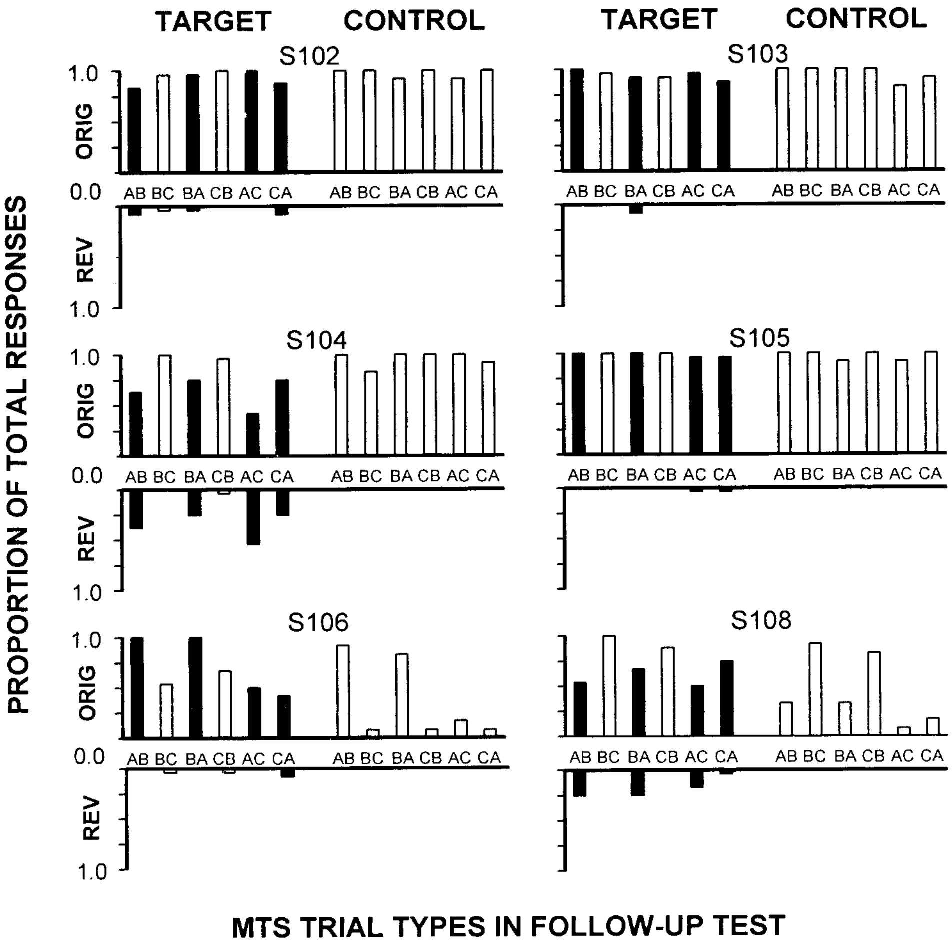

Fig. 5. Proportion of responses on each MTS trial type consistent with the original baseline training (ORIG) or AB reversal training (REV) aggregated across trial blocks of the follow-up test. All other details as in Figure 4.

tions, particularly semantic relations. For example, saying "car" in the presence of the words *Ford*, *Chevy*, and *Toyota* may be reinforced. Then, if saying "auto" is reinforced in the presence of the word *Ford*, both *Chevy* and *Toyota* also may set the occasion for saying "auto." Other studies have expanded functional classes through MTS procedures (e.g., Lazar, 1977) and have shown that respondent functions could be transferred through stimulus equivalence classes (e.g., Dougher, Augustson, Markham, Greenway, & Wulfert, 1994). We know of no studies, however, that have attempted to manipulate functional classes by reinforcing common responses in the presence of a set of arbitrarily related stimuli, reinforcing a new response in the presence of one stimulus, and testing whether this response transfers to other stimuli. The present experiment illustrates how this functional equivalence procedure can be used to establish, reverse, and measure the stability of stimulus classes.

In addition, the present study suggests a number of similarities between stimulus equivalence and functional equivalence that support Sidman's (1986, 1994) contention that the two types of class formation are controlled by similar selection processes. Some differences were found as well, but pending further empirical analysis, these may be attributed provisionally to unremarkable factors, such as asymmetries in the procedures used to establish the classes or generic differences in the selection processes that work on conditional discriminations versus simple discriminations.

Similarities Between Stimulus Equivalence and Functional Equivalence

Sensitivity to the reversal contingencies. In both conditions, only the two stimulus classes that were targeted for reversal were affected by the changes in baseline contingencies. The control classes in both conditions (i.e., A3B3C3) remained intact during the postreversal tests suggesting that simple-discrimination and MTS performances during the postreversal tests were in fact controlled by a manipulation of the class-defining contingencies.

Simple-discrimination performances were sensitive to reversal contingencies in a postreversal test. Five of 6 students reversed their responding to B and C stimuli after being trained to reverse responding to A stimuli. This outcome suggests that the discriminative functions of the original stimulus classes (i.e., A1B1C1 occasioned R1 and A2B2C2 occasioned R2) reversed, while stimulus class memberships remained intact (i.e., A1B1C1 occasioned R2 and A2B2C2 occasioned R1).

MTS performances also were sensitive to reversals of baseline contingencies. Acquisition of the reversed conditional discriminations required approximately the same number of trial blocks as acquisition of reversed simple discriminations. The AB baseline reversals resulted in corresponding reversals of MTS performances on BA symmetry, AC transitivity, and CA equivalence trial types for 4 of the 6 students, suggesting that the new stimulus equivalence classes A2B1C1 and A1B2C2 were established.

Such class reorganizations, however, were dependent on the reversal of the AB baseline relations. Initially, S102, S103, and S106 failed to reverse responding on the AB baseline relations. Not surprisingly, their performance on the subsequent postreversal test remained consistent with the original class organizations. For S102 and S106, a second round of AB reversal training and postreversal testing was sufficient to produce reversed responding and class reorganization. One student (S106), however, reversed symmetry trial types but only partially reversed transitivity and equivalence trial types. This dissociation among trial types is consistent with the findings of Pilgrim and Galizio (1990, 1995), who showed that transitivity and equivalence relations appear to be less sensitive to reversals of baseline contingencies than baseline and symmetry relations. For S103, however, even this second round of reversal training and testing failed to maintain reversed responding on the AB baseline relations during the postreversal test. This apparent insensitivity of MTS performances to the reversal contingencies is consistent with previous studies in which altering baseline relations sometimes was found to be quite difficult, especially when those changes conflict with organization of the original stimulus equivalence classes (e.g., Pilgrim et al., 1995; R. Saunders, Saunders, Kirby, & Spradlin, 1988).

Stability over time. Previous studies have found that stimulus equivalence classes, once established and demonstrated, can be quite stable over time (Rehfeldt & Hayes, 2000; R. Saunders, Wachter, & Spradlin, 1988; Spradlin et al., 1992). The present experiment replicated that finding, and apparently is the first to demonstrate similar stability with functional equivalence classes. This stability was evident even when class-defining contingencies were altered during the period between initial class formation and follow-up testing. The follow-up MTS and simple-discrimination performances were consistent mostly with the original baseline contingencies and not the most recently established altered classes. Thus, the effects of the reversal contingencies appeared to be temporary.

A likely contributor to the reinstatement of original patterns of responding for the stimulus equivalence and functional stimulus classes is the presence of a control class (e.g., A3B3C3) that remained unaltered. Thus, responding to the control classes consistent with original training may have strengthened original patterns of responding with respect to the other two stimulus classes.

Reemergence of original response patterns also may be explained in terms of differential reinforcement histories involving original baseline relations and reversed baseline relations. All baseline relations of the original stimulus classes were trained extensively. By contrast, the single reversed baseline relation was the object of much less training. Moreover, during reversal training, responding to the control classes, which were consistent with the original classes, was reinforced, thereby adding to the number of reinforcers provided for responding to the original classes. Separating these sources of control for responding to the original classes could be a topic for further investigations, as well as testing outcomes in the absence of the control classes.

Differences Between Stimulus Equivalence and Functional Equivalence

At a structural level, stimulus equivalence and functional equivalence encompass different behavior–environment relations, making it impossible to precisely equate the procedures used to instate, assess, and reverse them. Thus, for this reason, direct comparisons of quantitative measures such as accuracy must be undertaken with caution. The present experiment sought instead to promote a qualitative within-subject comparison of the effects of baseline reversals on simple- and conditional-discrimination performances associated with the two types of classes.

Differences in outcome. Differences were found during the initial tests for stimulus equivalence and functional equivalence. All students quickly exhibited class-consistent performances indicative of stimulus equivalence, most during the first trial block of each test. By contrast, functional equivalence emerged slowly. Only S108 demonstrated the establishment of a functional stimulus class immediately after the first function change, suggesting that A, B, and C stimuli initially had acquired independent discriminative functions (see also Smeets et al., 1995). This outcome was expected because the initial training conditions did not differentiate between individual discriminative functions and class functions. For most students, the formation of functional stimulus classes required that the transfer of the new functions be reinforced explicitly and then demonstrated in the absence of reinforcement after a second function change. Even following this training, 2 students failed the second test of transfer, further illustrating that even explicitly reinforcing class functions after one iteration of training and testing was not sufficient to produce class functions over independent discriminative functions for these students.

The effects of baseline reversals also were different. Following the reversal of a baseline relation, stimulus equivalence classes readily became reorganized to accommodate the new relation. By contrast, the reversal of selected baseline simple discriminations resulted not in a reorganization of the original functional stimulus classes, but in the complete reversal of function across the other class of stimuli. In other words, the organization of original functional stimulus classes remained intact after the reversal training, but the responses to these classes changed.

These outcomes are in accord with the reversal contingencies arranged in the present experiment. With stimulus equivalence classes, the reinforcement contingencies for class-defining conditional discriminations were altered, and therefore all relations derived from those prerequisite relations would be expected to change accordingly. With functional stimulus classes, however, the contingencies required reversing the responses instead of the class memberships. Once functional equivalence among sets of stimuli was established, by definition, any change in responding applied to one stimulus of a set should have applied similarly to the other stimuli of the set. The reinforcement contingencies during reversal training provided such a change, and the postreversal test allowed the demonstration of a transfer of function.

Most reports that have addressed the relation between stimulus equivalence and functional equivalence (e.g., Dube et al., 1991; Sidman et al., 1989; Vaughan, 1988) have been based on performances with functional stimulus classes that were established through repeated reversals of function. Although it is possible that the use of repeated reversals might have yielded more efficient demonstrations of functional equivalence, our test for functional equivalence was adopted in part to maintain some procedural consistency with the stimulus equivalence procedures. Training and testing for functional equivalence and stimulus equivalence were similar in that they both involved tests for novel performances. This followed from the argument (K. Saunders, Williams, & Spradlin, 1996) that using reversal of functions to establish functional stimulus classes does not result in novel or derived relations (i.e., untrained or never-before-seen behavior). For instance, tests for functional equivalence involving repeated reversals involve explicit reinforcement for transfer of function. Explicit reinforcement of transfer does not occur when novel responses are added to classes as in the current study. When responses are added, a generalized transfer of function may be tested, much as is the case with stimulus equivalence procedures. Thus, the second iteration of training and testing can be considered a test for generalized transfer, because new responses were trained and then tested in the absence of reinforcement.

Conclusions

Since Sidman (1986) and Vaughan (1988) originally proposed a relation between stimulus equivalence and functional equivalence, few studies have specifically addressed this relation. Toward that end, the present experiment employed procedures that have been used with MTS to study the establishment, reversal, and stability of arbitrary functional equivalence classes. Results of the present experiments suggest some potentially important similarities and differences between stimulus equivalence and functional equivalence.

Considerable evidence suggests that contingencies of reinforcement produce responding to classes of stimuli (Sidman, 2000). The contingencies of reinforcement that usually are involved in stimulus equivalence experiments select two kinds of responding: the analytic unit called conditional discrimination and the family of stimulus–stimulus relations called equivalence (Sidman, 2000). Whether subjects respond consistently with one of these outcomes or the other depends on other variables, including their history with classes of stimuli prior to the experiment and variations in the training and testing procedures. For example, when the training procedures do not ensure that individual discriminative relations among stimuli are stable, then reliable class-consistent responding does not occur (R. Saunders & Green, 1999). And when the testing conditions do not restrict responding to the relations used to test equivalence, classes do not form reliably (Duarte, Eikeseth, Rosales-Ruiz, & Baer, 1998; Innis, Lane, Miller, & Critchfield, 1998).

Similarly, the contingencies of reinforcement for functional equivalence select two kinds of responding: the analytic unit called simple discrimination and the family of stimulus–response relations called functional equivalence. As in stimulus equivalence, whether subjects respond to stimuli as a class depends on other variables. As a result, some tests for class membership permit multiple outcomes. For example, in the first testing phase of the present functional equivalence conditions, students could respond to the stimuli as individual simple discriminations or as members of a class. Only 1 student responded to stimuli as a class. When explicit contingencies required students to transfer and reverse functions, they did so readily. This seems similar to the finding with stimulus equivalence when the training and testing procedures are sufficiently restrictive to unambiguously select responding to classes of stimuli. Consistent with Sidman (2000), the formation of functional and stimulus equivalence classes is a product of the contingencies of reinforcement.

REFERENCES

Carlin, L. A., Wirth, O., & Chase, P. N. (1998). Effects of sample-response requirements on matching-to-sample performance with humans. *Experimental Analysis of Human Behavior Bulletin, 16,* 2–5.

Catania, A. C. (1998). The taxonomy of verbal behavior. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior: Applied clinical psychology* (pp. 405–433). New York: Plenum Press.

Cofer, C. N., & Foley, J. P. (1942). Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychological Review, 49,* 513–540.

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior, 62,* 331–351.

Duarte, A. M., Eikeseth, S., Rosales-Ruiz, J., & Baer, D. M. (1998). The effects of a can't-answer response option and instructions on stimulus equivalence. *The Psychological Record, 48,* 631–646.

Dube, W. V., McDonald, S. J., & McIlvane, W. J. (1991). A note on the relationship between equivalence classes and functional stimulus classes. *Experimental Analysis of Human Behavior Bulletin, 9,* 7–11.

Goldiamond, I. (1962). Perception. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology* (pp. 280–340). New York: Basic Books.

Goldiamond, I. (1966). Perception, language, and conceptualization rules. In B. Kleinmuntz (Ed.), *Problem solving* (pp. 183–224). New York: Wiley.

Goss, A. E. (1961). Verbal mediating responses and concept formation. *Psychological Review, 68,* 248–274.

Innis, A., Lane, S. D., Miller, E. R., & Critchfield, T. S. (1998). Stimulus equivalence: Effects of a default-response option on emergence of untrained stimulus relations. *Journal of the Experimental Analysis of Behavior, 70,* 87–102.

Lazar, R. (1977). Extending sequence-class membership with matching to sample. *Journal of the Experimental Analysis of Behavior, 27,* 381–392.
Peters, H. N. (1935). Mediate association. *Journal of Experimental Psychology, 18,* 20–48.
Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior, 63,* 239–254.
Pilgrim, C., & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal of the Experimental Analysis of Behavior, 54,* 213–224.
Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63,* 225–238.
Rehfeldt, R. A., & Hayes, L. J. (2000). The long-term retention of generalized equivalence classes. *The Psychological Record, 50,* 405–428.
Saunders, K. J., Williams, D. C., & Spradlin, J. E. (1996). Derived stimulus control: Are there differences among procedures and processes? In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 93–109). Amsterdam: Elsevier.
Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior, 72,* 117–137.
Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior, 50,* 145–162.
Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 49,* 95–115.
Sidman, M. (1986). Functional analysis of emergent stimulus classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.
Sidman, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors Cooperative.
Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74,* 127–146.
Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discrimination of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior, 37,* 23–44.
Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.
Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52,* 261–274.
Smeets, P. M. (1994). Stability of emergent simple discriminations in young children. *Journal of Experimental Child Psychology, 57,* 397–417.
Smeets, P. M., Barnes, D., & Luciano, M. C. (1995). Reversal of emergent simple discrimination in children: A component analysis. *Journal of Experimental Child Psychology, 60,* 327–343.
Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 29–42). Reno, NV: Context Press.
Touchette, P. E. (1971). Transfer of stimulus control: Measuring the moment of transfer. *Journal of the Experimental Analysis of Behavior, 15,* 347–354.
Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 36–42.
Wirth, O., Chase, P. N., & Munson, K. J. (2000). Experimental analysis of human vocal behavior: Applications of speech-recognition technology. *Journal of the Experimental Analysis of Behavior, 74,* 363–375.
Zentall, T. R., & Smeets, P. M. (Eds.). (1996). *Stimulus class formation in humans and animals.* Amsterdam: Elsevier.

Received March 14, 2001
Final acceptance September 11, 2001

APPENDIX A

Experimental phases and composition of trial blocks for reversals with functional stimulus classes. Italicized responses indicate reinforced responses during baseline and reversal training. Students' responses were restricted to those shown for training phases; however, students could emit any previously trained response during functional equivalence and postreversal tests.

	Phase	Trials of each type/total per block	Trial types (discriminative stimulus: responses) Target classes	Control class
1.	Original baseline training			
	AR	8/24	A1: *R1,* R2, R3	A3: R1, R2, *R3*
			A2: R1, *R2,* R3	
	BR	8/24	B1: *R1,* R2, R3	B3: R1, R2, *R3*
			B2: R1, *R2,* R3	
	CR	8/24	C1: *R1,* R2, R3	C3: R1, R2, *R3*
			C2: R1, *R2,* R3	
	AR, BR, CR mix	4/36	A1: *R1,* R2, R3	A3: R1, R2, *R3*
			A2: R1, *R2,* R3	B3: R1, R2, *R3*
			B1: *R1,* R2, R3	C3: R1, R2, *R3*
			B2: R1, *R2,* R3	
			C1: *R1,* R2, R3	
			C2: R1, *R2,* R3	
2.	Test of functional equivalence			
	Function change	8/24	A1: *R4,* R5, R6	A3: R4, R5, *R6*
			A2: R4, *R5,* R6	
	Transfer test	4/36	A1: R4, R5, R6	A3: R4, R5, R6
			A2: R4, R5, R6	B3: R4, R5, R6
			B1: R4, R5, R6	C3: R4, R5, R6
			B2: R4, R5, R6	
			C1: R4, R5, R6	
			C2: R4, R5, R6	
3.	Baseline retraining	4/36	(same as AR, BR, CR mix above)	
4.	Reversal training (AR only)	8/24	A1: R1, *R2,* R3	A3: R1, R2, *R3*
			A2: *R1,* R2, R3	
5.	Postreversal test (and follow-up test)	4/36	A1: R1, R2, R3	A3: R1, R2, R3
			A2: R1, R2, R3	B3: R1, R2, R3
			B1: R1, R2, R3	C3: R1, R2, R3
			B2: R1, R2, R3	
			C1: R1, R2, R3	
			C2: R1, R2, R3	

APPENDIX B

Experimental phases and composition of trial blocks for reversal with stimulus equivalence classes. Italicized comparison stimuli indicate reinforced selections during baseline and reversal training. Equivalence and postreversal tests included AB and BC baseline trial types interspersed among the test trial types.

	Phase	Trials of each type/total per block	Trial types (sample: comparisons) Target classes	Control class
1.	Original baseline training			
	AB	8/24	A1: *B1,* B2, B3	A3: B1, B2, *B3*
			A2: B1, *B2,* B3	
	BC	8/24	B1: *C1,* C2, C3	B3: C1, C2, *C3*
			B2: C1, *C2,* C3	
	AB, BC mix	4/24	A1: *B1,* B2, B3	A3: B1, B2, *B3*
			A2: B1, *B2,* B3	B3: C1, C2, *C3*
			B1: *C1,* C2, C3	
			B2: C1, *C2,* C3	
2.	Tests of stimulus equivalence			
	Reflexivity (AA, BB, CC)	4/60	A1: A1, A2, A3	A3: A1, A2, A3
			A2: A1, A2, A3	B3: B1, B2, B3
			B1: B1, B2, B3	C3: C1, C2, C3
			B2: B1, B2, B3	
			C1: C1, C2, C3	
			C2: C1, C2, C3	
	Symmetry (BA, CB)	4/48	B1: A1, A2, A3	B3: A1, A2, A3
			B2: A1, A2, A3	C3: B1, B2, B3
			C1: B1, B2, B3	
			C2: B1, B2, B3	
	Transitivity (AC)	4/36	A1: C1, C2, C3	A3: C1, C2, C3
			A2: C1, C2, C3	
	Equivalence (CA)	4/36	C1: A1, A2, A3	C3: A1, A2, A3
			C2: A1, A2, A3	
3.	Baseline reversal (AB only)	8/24	A1: B1, *B2,* B3	A3: B1, B2, *B3*
			A2: *B1,* B2, B3	
4.	Postreversal test (and follow-up test)	4/72	B1: A1, A2, A3	B3: A1, A2, A3
			B2: A1, A2, A3	C3: B1, B2, B3
			C1: B1, B2, B3	A3: C1, C2, C3
			C2: B1, B2, B3	C3: A1, A2, A3
			A1: C1, C2, C3	
			A2: C1, C2, C3	
			C1: A1, A2, A3	
			C2: A1, A2, A3	